

09/669,680

REMARKS

Claims 1, 3-8, 10-23, and 25-29 are all the claims pending in the application. Claims 2, 9, and 24 have been incorporated into independent claims 1, 8, and 23, respectively. Claims 1, 3-8, 10-23, and 25-29 stand rejected on prior art grounds. Applicants respectfully traverse this rejection based on the following discussion.

I. The Prior Art Rejections

Claims 1-29 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Lantrip et al., hereinafter "Lantrip" (U.S. Patent No. 6,298,174) in view of Ruocco et al., hereinafter "Ruocco" (U.S. Patent No. 5,864,855). Applicants respectfully traverse these rejections based on the following discussion.

A. The Rejection Based on Lantrip in view of Ruocco

In paragraph 34 of the Office Action, the Office argues that the combination of Ruocco and Lantrip teaches clustering that is based on prior categorization efforts. However, Applicant maintains the previously presented arguments that neither reference teaches a second dataset. As discussed previously, Lantrip is only concerned with the formation of a single dataset. Ruocco clusters documents into a single dataset in a process that simultaneously compares a new document that is being added to the dataset with previously established clustering centroids to evaluate in which previously established cluster the new document should be added. While Ruocco uses the previously established clustering centroids when adding the new document to the previous clusters, Ruocco does not cluster "a new, but related dataset different than said first dataset" using the previously established clustering centroids as defined by independent claims 1, 8, 15, 20, and 23. Instead, Ruocco merely enlarges the previous dataset and does not cluster a new dataset that is different from the previous dataset.

09/669,680

The clustering described in Ruocco keeps all the old legacy clusters around, even if they are no longer relevant. The claimed method starts with the old cluster definitions (the centroids generated from the first data set), but lets those centroids shift immediately to fit the data in the second data set, ignoring completely any data in the first data set, thereby allowing old concepts to disappear and completely new ones to emerge more readily.

Applicants respectfully traverse this rejection because the applied prior art references do not teach or suggest "clustering second documents in a second dataset using said centroid seeds, such that said second dataset has a similar clustering to that of said first dataset" as defined by independent claims 1, 20, and 23; "said cluster generator clusters second documents in said second dataset using said centroid seeds, such that said second dataset has a similar clustering to that of said first dataset" as defined by independent claim 8 or "clustering said second documents using said centroid seeds, such that said second dataset has a similar clustering to that of said first dataset" as defined by independent claim 15.

The claimed invention solves the problem of finding new categories in a second data set that did not exist in the first data set, while at the same time maintaining as nearly as possible categories from the first data set as categories in the second data set. With the claimed invention, there is no requirement, and in fact it is not assumed, that the first and second data sets have any of the same data elements in them. They are allowed to have some of the same elements, but this is in no way a requirement for the claimed invention. The claimed invention is designed in such a way as to find the similarities between the two data sets, where they exist, while at the same time finding the key differences (emerging concepts) in the second data set.

Therefore, Applicants submit that the applied prior art references do not teach or suggest "clustering second documents in a second dataset using said centroid seeds, such that said second dataset has a similar clustering to that of said first dataset" as defined by independent claims 1, 20, and 23; "said cluster generator clusters second documents in said second dataset using said centroid seeds, such that said second dataset has a similar clustering to that of said first dataset" as defined by independent claim 8 and "clustering said second documents using said centroid seeds, such that said second dataset has a similar clustering to that of said first dataset" as defined

09/669,680

by independent claim 15. Therefore, independent claims 1, 8, 15, 20, and 23 are patentable over the applied prior art references. Further, dependent claims 3-7, 10-14, 16-19, 21, 22, and 25-29 are similarly patentable, not only by virtue of their dependency from a patentable claim, but also by virtue of the additional features of the invention they define. In view of the foregoing, the Examiner is respectfully requested to reconsider and withdraw this rejection.

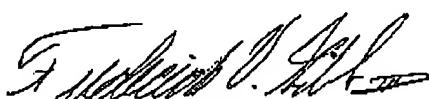
II. Formal Matters and Conclusion

In view of the foregoing, Applicant submits that claims 1, 3-8, 10-23, and 25-29, all the claims presently pending in the application, are patentably distinct from the prior art of record and are in condition for allowance. The Examiner is respectfully requested to pass the above application to issue at the earliest possible time.

Please charge any deficiencies and credit any overpayments to Attorney's Deposit
Account Number 09-0441.

Respectfully submitted,

Dated: 1-13-05



Frederick W. Gibb, III
Reg. No. 37,629

McGinn & Gibb, PLLC
2568-A Riva Road
Suite 304
Annapolis, MD 21401
Customer Number: 29154